



Repérage de relations sémantiques entre termes : sur la piste de la morphologie dérivationnelle

Natalia Grabar, Thierry Hamon

► To cite this version:

Natalia Grabar, Thierry Hamon. Repérage de relations sémantiques entre termes : sur la piste de la morphologie dérivationnelle. 15èmes journées francophones d'Ingénierie des Connaissances, May 2004, Lyon, France. pp.237-248. hal-00084834

HAL Id: hal-00084834

<https://hal.science/hal-00084834>

Submitted on 16 Mar 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Repérage de relations sémantiques entre termes : sur la piste de la morphologie dérivationnelle

Natalia Grabar¹, Thierry Hamon²

¹ STIM-DSI, AP-HP ; ERM 202 INSERM
Département de Biomathématiques, Université Paris 6
`ngr@biomath.jussieu.fr`

² LIPN – UMR CNRS 7030
Université Paris-Nord
Avenue J.B. Clément, 93430 Villetaneuse, FRANCE
`thierry.hamon@lipn.univ-paris13.fr`

Résumé :

Notre travail est consacré au repérage de relations sémantiques entre termes. Dans le contexte de constitution de terminologies structurées, nous nous intéressons en particulier à l'aide que peut apporter une approche basée sur la morphologie au regard d'autres techniques d'acquisition de relations sémantiques. Parmi les opérations dont dispose la morphologie, nous nous concentrons sur l'affixation. En fonction des opérateurs morphologiques, nous montrons quelques schémas interprétatifs qui se dégagent de l'analyse des termes et indiquons les relations sémantiques qui sont alors aptes à émerger.

Mots-clés : Terminologie, structuration de termes, morphologie dérivationnelle.

1 Introduction

Les produits terminologiques ont vocation à décrire la connaissance d'un secteur d'activité. Cette description peut être plus ou moins détaillée et fine : elle va d'une liste de termes « à plat » à un réseau de termes structurés sémantiquement. Nous nous intéressons ici à la détection de relations sémantiques entre termes à l'aide de la morphologie, et en particulier à l'aide de la morphologie dérivationnelle. Cette piste peut s'avérer pertinente dans les domaines où la morphologie est particulièrement sollicitée lors de la création des termes, comme le font la médecine, la biologie, l'agronomie, etc. Les expériences présentées ci-dessous militent en faveur de l'utilisation de la morphologie : au même titre que d'autres

types d'approches, les indices morphologiques sont exploitables pour relier sémantiquement des termes, mais aussi pour préciser les relations déjà établies. Nous présentons donc les pistes proposées par la morphologie, essentiellement au moyen de l'affixation, pouvant contribuer à la structuration des termes.

Nous présentons tout d'abord les types de relations pouvant exister dans une terminologie structurée (sec. 2), et les approches permettant la structuration des termes (sec. 3). Puis, nous présentons les indices proposés par la morphologie constructionnelle, particulièrement dérivationnelle, (sec. 4), exploités pour mettre en relation des termes issus de deux domaines : la médecine et la cogénération (sec. 5). Nous présentons et discutons les résultats obtenus à partir de ces pistes (sec. 6).

2 Relations dans une terminologie structurée

La structuration d'une liste de terme peut être réalisée à l'aide de plusieurs types de relations sémantiques. Selon la nature des termes reliés, nous distinguons les relations suivantes (Grabar & Hamon, 2004) :

- Les *relations taxinomiques* organisent des concepts dans une arborescence.

La *relation hiérarchique* (ou *est-un*) relie un terme générique à ses termes plus spécifiques. Cette relation, la plus présente dans les produits terminologiques, en constitue, traditionnellement, l'épine dorsale. Les exemples ci-dessous sont issus de la terminologie médicale SNOMED (Côté, 1996) :

| | | |
|-----------------------------------|---------------|---|
| <i>pneumonie</i> | <i>est-un</i> | <i>bronchopneumonie</i> |
| <i>pneumopathie inflammatoire</i> | <i>est-un</i> | <i>bronchopneumonie</i> |
| <i>bronchopneumonie</i> | <i>est-un</i> | <i>maladie de l'appareil respiratoire</i> |

La *relation partitive* (*méronymie* ou *partie-tout*), souvent utilisée pour décrire les artefacts et les organismes vivants à travers l'énumération de leurs parties constituantes, peut également assurer ce rôle au côté de la relation hiérarchique :

poumon partie-de appareil respiratoire

- Les *relations lexicales* structurent les termes possédant un même hyperonyme. Elles regroupent deux types de relations apparaissant moins fréquemment dans les produits terminologiques.

La *relation d'équivalence*, ou de *synonymie*, rassemble les termes qui réfèrent à un même concept ¹. Par exemple, les termes *pneumonie* et *pneumopathie inflammatoire* sont synonymes dans la terminologie médicale SNOMED. La synonymie permet de compléter et de préciser un noeud conceptuel. Elle peut également relier les variantes d'un terme.

La *relation d'opposition*, de *contraire* ou d'*antonymie* relie les co-hyponymes qui ne sont pas des synonymes (Amsili, 2003). Cette relation est

1. Dans l'acception que nous en faisons ici, les *termes* correspondent aux expressions linguistiques telles qu'on les retrouve dans les documents. Il s'agit d'unités textuelles ou linguistiques. Les *concepts* correspondent à des notions abstraites et existent au niveau conceptuel.

définie par exemple dans la terminologie de la cogénération, entre les termes *énergie nucléaire* et *énergie non nucléaire* (Grabar & Jeannin, 2002).

- Les *relations transversales* relient les concepts situés dans des branches différentes de la hiérarchie. Ces relations, souvent annotées comme relation **voir-aussi**, sont majoritairement sous-spécifiées alors qu’elles conduisent à une description plus fine des connaissances du domaine. De plus, dans les terminologies *post-coordonnées*, elles assurent la fonction définitoire des concepts en permettant la constitution de nouveaux concepts à partir de concepts « élémentaires » déjà recensés et de règles de compatibilité et de composition sémantique (Spackman & Campbell, 1998; Zweigenbaum, 1999). Par exemple, le diagnostic *pneumonie* est défini comme une atteinte morphologique de type *inflammation* et se trouve localisé dans une partie du corps (*poumon*) :

$$\begin{array}{lcl} \textit{pneumonie} & \rightarrow & \text{est-un} \rightarrow \textit{inflammation} \\ & \searrow & \text{localisé-dans} \rightarrow \textit{poumon} \end{array}$$

Dans la section suivante, nous présentons les travaux de repérage (semi)automatique de relations sémantiques entre termes et précisons les types de relations pouvant alors être détectés.

3 Approches pour l’identification de relations sémantiques entre termes

Nous distinguons deux grands types d’approches pour la détection de relations sémantiques entre termes (Grabar & Hamon, 2004) : les approches basées sur l’étude du contexte des termes (sec. 3.1) et les approches basées sur l’analyse de la structure interne des termes (sec. 3.2). Ces approches se focalisent sur un type de relation et s’appuient sur une expertise humaine finale.

Les relations habituellement recherchées sont principalement des relations taxinomiques, alors que la synonymie l’est plus rarement malgré son importance dans les terminologies. L’acquisition de relations transversales est, quant à elle, négligée mais profite de l’acquisition de deux autres types de relations.

3.1 Étude des contextes des termes

L’étude du contexte peut avoir pour objectif de repérer, dans un corpus textuel, les expressions ou les phrases susceptibles de contenir un certain type de relations entre deux termes. Il s’agit alors d’utiliser les patrons lexico-syntaxiques (Hearst, 1992; Séguéla & Aussenac-Gilles, 1999; Morin, 1999) ou des marqueurs caractéristiques des relations recherchées (Garcia, 1997; Pearson, 1998). Par exemple, les patrons :

```
such NP as {NP ,}* {or|and} NP
NP {,} including {NP ,}* {or|and} NP
```

permettent de repérer les relations hiérarchiques (Hearst, 1992).

Il est également possible d'aborder l'étude des contextes de termes sans avoir des *a priori* quelconques sur les relations à découvrir en corpus. Les approches de type distributionnel visent à repérer les contextes terminologiques communs. Moyennant une interprétation des résultats par un expert du domaine, ces termes peuvent être regroupés en classes (Grefenstette, 1994; Nazarenko *et al.*, 2001; Assadi, 1998). Ainsi, (Grefenstette, 1994) interprète les classes de mots :

large : *small, major, great, various, main, different, field*

patient : *case, group, child, day, treatment, woman*

pour identifier des relations d'antonymie (*large, small*), de synonymie (*large, great*), de méronymie (*patient, group*), d'hyperonymie (*patient, child, woman*) ou bien encore la relation transversale (*patient, treatment*).

3.2 Étude de la structure des termes

L'analyse de la structure interne des termes permet également l'acquisition des relations. Ainsi l'hypothèse d'inclusion lexicale conduit à supposer que si un terme est inclus dans un autre, alors il existe une relation sémantique entre eux (Bourigault, 1993; Bodenreider *et al.*, 2001; Grabar & Zweigenbaum, 2003). De nombreux types de relations peuvent être identifiés :

relation hiérarchique : *électricité nucléaire est-un électricité*

antonymie : *électricité nucléaire anto électricité non nucléaire*

relation transversale : *station électrique produit électricité*

Le repérage de synonymes et de variantes morpho-syntaxiques des termes s'appuient plutôt sur des approches à base de règles de transformation. (Jacquemin, 1996) décrit ainsi la variation morpho-syntaxique : par exemple, la règle $N_1N_2 \rightarrow N_1 \text{ or } N_3N_2$ définit la coordination de termes (*X ray diffraction / X ray or neutron diffraction*). De même, pour acquérir des relations de synonymie entre termes complexes, (Hamon *et al.*, 1998) utilisent la propagation compositionnelle : deux termes complexes sont considérés comme synonymes si leurs composants sont identiques ou synonymes (*protection syno sauvegarde* \rightarrow *action de protection syno action de sauvegarde*).

Le travail présenté dans cet article est basé également sur l'étude de la forme interne des termes pour identifier des relations. Alors que les travaux décrits ci-dessus exploitent des indices principalement syntaxiques, nous nous intéressons ici à l'analyse de la structure interne des termes appuyée par la morphologie.

4 Morphologie constructionnelle

Nous présentons maintenant les opérations dont dispose la morphologie constructionnelle et que nous supposons être utiles pour la structuration des termes. La morphologie constructionnelle (Corbin, 1987) stipule que le sens des mots est construit en même temps que leur forme. Ainsi, les règles de formation des mots encodent cette information à travers les contraintes d'ordre phonologique,

syntactique, sémantique et morphologique. La morphologie constructionnelle dispose de plusieurs types d'opérations morphologiques décrites à travers les règles de formation des mots :

- L'*affixation*, ou la *dérivation*, combine une base (*artère*) avec un affixe, suffixe ou préfixe, pour former un mot affixé (*artériel*).
- La *conversion* décrit les cas de passage d'un mot d'une catégorie syntaxique à une autre: {*muqueuse*/A, *muqueuse*/N}.
- La *composition* combine au moins deux bases pour former un mot composé: *artère* et *scopie* forment *artérioscopie*.

De plus, bien que la *supplétion* ne soit pas un type d'opération morphologique mais un moyen de mise en correspondance de bases équivalentes, elle participe à l'affixation comme à la composition en mettant en relation des bases équivalentes qui proviennent de différentes langues, notamment grecque et latine. Par exemple, *gastr-* (base d'origine grecque) est substitué à *estomac* (base d'origine latine) lors de la formation de l'adjectif dénominal *gastrique*.

Les opérations d'affixation constituent les opérations de base dans la formation des lexèmes et sont donc présentes dans toutes les langues de spécialité. La composition est liée à l'utilisation de bases d'origine latine ou grecque et semble, de ce fait, être réservée à certaines des langues de spécialité. La supplétion, comme la composition, va de paire avec l'utilisation des bases d'origine grecque et latine. Elle pourrait aussi être marquée par certaines langues de spécialité. L'affixation, la composition et la supplétion peuvent conduire au repérage de différentes relations sémantiques entre termes. La nature globale de ces relations peut être cernée à travers les instructions sémantiques des règles de formation des lexèmes. Dans la suite de cet article, nous nous concentrons sur les pistes proposées par l'affixation.

5 Méthodes et matériel

Les méthodes utilisées visent à induire des relations sémantiques entre termes ou bien à affiner les résultats de structuration obtenus avec d'autres approches. Nous affinons ainsi les relations induites avec la méthode décrite dans (Grabar & Zweigenbaum, 2003) où nous avons appliqué l'hypothèse de l'inclusion lexicale pour le repérage de relations hiérarchiques. Par contre, l'observation de ces résultats a montré que des relations transversales sont également proposées. Nous nous basons sur la suffixation pour mieux cerner ces relations transversales. Une modification de la méthode (Grabar & Zweigenbaum, 2003) nous permet d'identifier des termes en relations d'opposition et en relation de localisation relative. Les relations qui marquent l'agent, l'action et le nom collectif sont repérées par rapport aux suffixes correspondants. Les expériences sont réalisées avec 11 586 termes provenant de la terminologie médicale SNOMED (Côté, 1996) et 4 117 termes de la cogénération (Grabar & Jeannin, 2002).

Dans les expériences présentées ici, nous exploitons les *règles de formation de lexèmes et leurs instructions sémantiques* pour détecter les relations transversales entre termes. Dans la section suivante, les règles en question et leur instruction sémantique sont décrites, puis nous discutons les termes mis en relation au moyen de ces règles. Notre approche nécessite des *ressources morphologiques*. Nous utilisons celles obtenues avec les méthodes décrites dans (Grabar & Zweigenbaum, 2000) et appliquées au domaine médical. Toutefois, l’affixation, même étant un moyen universel de formation de lexèmes, demande des ressources morphologiques spécifiques à chacune de ces langues (Grabar & Zweigenbaum, 2003). Les ressources morphologiques, utilisées avec les termes de la cogénération, y ont donc été adaptées.

Dans le domaine médical, pour mieux cerner et vérifier la sémantique des termes lors de l’analyse des résultats, nous nous appuyons également sur un *dictionnaire médical* (Manuila *et al.*, 2001).

6 Les expériences et la discussion des résultats

L’affixation permet de former des mots au moyen de préfixes et de suffixes. Le sens des mots affixés est le résultat de l’opération sémantique qu’exerce l’affixe sur le sens de la base. Nous présentons ici les expériences réalisées avec certaines opérations de préfixation (sec. 6.1) et de suffixation (sec. 6.2). Pour chaque opération impliquée dans les expériences, nous donnons son instruction sémantique et discutons les résultats de son application.

6.1 Préfixation

Construction de sens contraire

L’antonymie peut être construite, par exemple, avec des préfixes *dé-*, *ir-*, *anti-*, *non-* et *in-* négatifs, ou bien avec *a-* et *dys-* privatifs (Amiot, 1998). L’utilité de certains de ces préfixes pour la structuration des termes est signalée par (Daille, 2003). Nous avons utilisé un ensemble de 52 paires de mots reliés avec de tels préfixes (*{ accessible, inaccessible }*, *{ fonction, dysfonction }*), ce qui nous a permis de relier 40 paires de termes médicaux, dont 30 termes complexes :

{ activateur du plasminogène, inactivateur du plasminogène },
{ kératose, dyskératose },
{ continence fécale, incontinence fécale }.

Toutes les propositions sont jugées être correctes.

Construction de sens de localisation spacia

Les relations transversales de localisation spatiale relative peuvent être construites avec les préfixes comme *sur-*, *sous-*, *contre-*, *péri-*, etc. (Amiot, 1998). Nous avons utilisé 99 paires de mots reliés grâce à 12 préfixes. Nous avons ainsi pu relier 40 paires de termes médicaux dont 21 termes complexes :

{*abcès rénal*, *abcès périrénal*},
{*hyperplasie kystique*, *hyperplasie intrakystique*},
{*cervicite chronique*, *endocervicite chronique*}.

Là aussi, toutes les propositions semblent être correctes.

D'autres préfixes peuvent conduire à d'autres relations éventuellement utiles dans une terminologie structurée : localisation temporelle (*pré-*, *post-*), comparaison (*super-*), etc.

6.2 Suffixation

Construction des noms d'agent et d'action

Les règles de construction des noms d'agent et d'action appliquent toutes les deux des suffixes aux verbes. Le nom d'agent est formé avec le suffixe *-eur* et porte l'instruction sémantique *Celui qui V*. Les noms d'action ou de résultat d'une action sont formés au moyen d'un ensemble de suffixes (*-age*, *-ade*, *-erie*, *-ment*, *-tion* et *-ure*) et portent l'instruction sémantique *Action de V* ou *Résultat de V*. Les deux règles construisent des relations transversales, qui marquent l'action, son résultat ou son acteur (Daille, 2003). Les verbes ou les groupes verbaux, par rapport auxquels ces relations sémantiques sont établies, sont rarement présents dans les terminologies. Les termes marqués *action*, *agent* ou *résultat d'une action* doivent donc être articulés autrement.

Le suffixe *-eur*, bien qu'ambigu, conduit assurément au repérage de termes signifiant les agents d'une action :

Médecine : *activateur du plasmogène*, *buveur modéré de boisson alcoolisée*, *gros fumeur de cigarettes (plus de 20 cigarettes par jour)*, *marqueur lymphocytaire*, *dialyseur péritonéal* ;
Cogénération : *producteur d'électricité*, *capteur solaire*, *consommeuur éligible*, *disjoncteur de couplage*, *cogénérateur*, *constructeur de turbine*, *compresseur à gaz*.

Il en est de même avec les suffixes qui marquent l'action ou son résultat :

Médecine : *blessure par balle*, *brûlure avec carbonisation*, *blocage congénital*, *tamponade*, *tatouage*, *ablation*, *absorption intestinale anormale* ;
Cogénération : *alimentation électrique*, *production éolienne*, *raccordement au réseau*, *aspiration*, *refoulement*, *protection électrique*, *reformage*, *recouplage*.

Construction des noms au sens collectif

La règle de construction de noms au sens collectif utilise par exemple les suffixes *-ail(le)*, *-ade*, *-age* ou *-ure* (Aliquot-Suengas, 1996). Le nom construit signifie qu'il comporte une ou plusieurs occurrences du nom de base et conduit potentiellement à la relation de méronymie (Aliquot-Suengas, 1996; Daille, 2003) :

Médecine : *ossature*, *denture*, *palmature congénitale*, *vergeture*, *vomissure gastrique*, *arcade sus-pyramidale du rein*, *cordage tendineux* ;
Cogénération : *sciure*, *outillage*, *câblage électrique*.

Construction d'adjectifs dénominaux

La règle de formation des adjectifs dénominaux, ou relationnels, applique aux noms un ensemble de suffixes, par exemple *-aire*, *-el*, *-al*, *-ique*, *-eux*, *-ien*, *-in*, *-ois* et *-é*. Les adjectifs construits reçoivent l'instruction sémantique *Relatif à N* et permettent donc l'identification de relations sémantiques entre le nom de base et l'adjectif qui en dérive: { *aorte*/Nom, *aortique*/Adj }.

Mais les adjectifs relationnels établissent également une relation entre leur nom de base (*aorte*) et leur nom recteur (*sténose*) (Mélis-Puchulu, 1991) dans les syntagmes nominaux où ils apparaissent (*sténose aortique*). Cette information peut être exploitée pour le repérage de relations sémantiques entre un terme simple ou complexe et un autre terme qui comporte l'adjectif obtenu par suffixation à partir du nom recteur du premier terme. (Mélis-Puchulu, 1991) distingue deux grands types de relation entre l'adjectif relationnel et son nom recteur, qui mènent aux relations de méronymie: l'appartenance et la possession.

- La relation d'appartenance est construite avec les suffixes *-é*, *-aire*, *-eux*, *-in* et *-ique*. Dans cette relation, le nom recteur signifie le tout et le nom de base de l'adjectif la partie:

nerf denté: *Nerf en forme des dents, Nerf qui a des dents.*

- La relation de possession est construite avec les suffixes *-al*, *-aire*, *-el*, *-ien*, *-in*, *-ique* et *-ois*. Dans cette relation, le nom recteur signifie la partie et le nom de base de l'adjectif le tout:

nerf dental: *Nerf qui se trouve dans une dent.*

Dans ces formations, à côté des suffixes non ambigus (*-é* et *-eux* pour l'appartenance, *-al*, *-el*, *-ien* et *-ois* pour la possession), les suffixes *-aire*, *-in* et *-ique* peuvent apparaître dans les formations des deux types.

Appliquée aux termes médicaux et à ceux de la cogénération, la relation existant entre l'adjectif et son nom de base, permet de détecter de nombreux couples de termes. Notons que les relations sémantiques sont plus fiables si le premier terme consiste en un nom et le deuxième en un groupe nominal simple, comme dans les exemples qui suivent. Les dépendances syntaxiques sont alors plus faciles à gérer. Nous analysons d'abord quelques exemples de termes médicaux:

| | | | |
|---------------------|---|----------------------------------|---------------|
| <i>abdomen</i> | ⇒ | <i>abcès abdominal</i> | possession, |
| <i>amygdale</i> | ⇒ | <i>noyau amygdalien</i> | possession, |
| <i>anévrisme</i> | ⇒ | <i>hématome anévrisimal</i> | possession, |
| <i>artère</i> | ⇒ | <i>cône artériel</i> | possession, |
| <i>achromie</i> | ⇒ | <i>mélanome achromique</i> | appartenance, |
| <i>actinomycose</i> | ⇒ | <i>infection actinomycosique</i> | appartenance, |
| <i>athérome</i> | ⇒ | <i>embolie athéromateuse</i> | appartenance. |

Ces exemples montrent que l'instruction sémantique globale (appartenance et possession) peut recevoir dans les langues de spécialité un sens plus précis. Dans les cas de possession et lorsque (1) le nom de base de l'adjectif signifie une partie du corps ou un tissu (*T*, topologie: *abdomen*, *amygdale*, etc.) et (2) le nom recteur de l'adjectif signifie une atteinte morphologique (*M*, maladie: *abcès*, *fibrosarcome*, etc.) la relation correspond à une localisation:

abcès abdominal localisé-dans *abdomen*,
noyau amygdalien localisé-dans *amygdale*

Mais les cas de possession offrent également d'autres types de relations :

cône artériel conduit-vers *artère*,
hématome anévrisimal provoqué-par *anévrisme*.

Le premier exemple semble présenter le schéma qui combine deux topologies ($T + T$), le deuxième deux morphologies ($M + M$). *Cirrhose alcoolique* et *acido-cétose alcoolique*, où nous n'avons pas pu définir s'il s'agit de la possession ou de l'appartenance, sont provoqué-par *alcool*. Dans ces exemples, une atteinte morphologique M est provoqué-par un produit chimique C . Les cas d'appartenance présentent le schéma déjà vu $M + M$, où la deuxième a provoqué la première.

Des cas plus complexes se présentent lorsqu'un terme comporte plus d'un adjectif relationnel, qui sont alors en relation de possession et d'appartenance. Les schémas sémantiques sont alors également plus complexes. Voici quelques termes avec deux adjectifs relationnels :

ganglion lymphatique abdominal : *ganglion* contient *lymphe* localisé-dans *abdomen* ($M + T + T$),
vésicule cutanée acantholytique : *vésicule* localisé-dans *peau* provoqué-par *acantholyse* ($M + T + M$),
maladie vasculaire artériosclérotique : *maladie* localisé-dans *vaisseau* provoqué-par *artériosclérose* ($M + T + M$).

L'analyse de la structure interne des termes médicaux qui comportent les adjectifs relationnels et l'établissement de leur schémas sémantiques a été facilité par l'existence de types sémantiques bien établis des termes. Il serait intéressant de comparer ces schémas avec ceux qui émergent des termes morphologiquement composés, comme *pyosalpingo-oophorite*, c'est-à-dire *inflammation de la trompe de Fallope et de l'ovaire avec formation de pus* ($M + T + T + \text{ite} (M)$) (Pacak et al., 1980). Notons cependant que le recteur syntaxique se trouve à gauche dans les constructions syntaxiques et à droite dans les constructions de la composition morphologique savante. Par ailleurs, dans le souci d'homogénéité du matériel, les analyses morpho-sémantiques des composés morphologiques sont faites par rapport au morphème recteur (*-ite*, *-ose*, *-ectomie*, etc.).

Nous présentons et analysons maintenant quelques exemples du domaine de la cogénération :

| | | | |
|--------------------|---|--------------------------------|---------------|
| <i>atmosphère</i> | ⇒ | <i>polluant atmosphérique</i> | possession, |
| <i>industrie</i> | ⇒ | <i>déchet industriel</i> | possession, |
| <i>troposphère</i> | ⇒ | <i>ozone troposphérique</i> | possession, |
| <i>gaz</i> | ⇒ | <i>acide gazeux</i> | appartenance, |
| <i>métallurgie</i> | ⇒ | <i>industrie métallurgique</i> | appartenance, |
| <i>carbone</i> | ⇒ | <i>gaz carbonique</i> | appartenance, |
| <i>soufre</i> | ⇒ | <i>rejet soufré</i> | appartenance. |

Dans les cas de possession, lorsque le nom recteur signifie une substance chimique (*polluant*, *ozone*), il existe une relation de localisation :

polluant atmosphérique localisé-dans *atmosphère*,
ozone troposphérique localisé-dans *troposphère*.

Par contre, un *déchet industriel* est **produit-par** l'*industrie*. Dans les cas d'appartenance, lorsque le nom de base de l'adjectif signifie un produit chimique (*carbone*, *soufre*), il s'agit de la relation méronymique. Les deux autres exemples (*acide gazeux* et *industrie métallurgique*) sont plus difficiles à caractériser. Un *acide gazeux* existe sous forme de *gaz*. L'instruction sémantique de base du suffixe *-eux* semble pleinement satisfaire la relation entre les deux termes. Et enfin l'*industrie métallurgique*, où, par rapport au terme *métallurgie*, il existe peu de changement sémantique.

De tous les exemples générés et présentés, il apparaît que l'affixation peut conduire au repérage de différents types de relations sémantiques entre les termes : la méronymie, l'antonymie et de nombreuses relations transversales. Nous n'avons pas analysé ici les relations d'équivalence pouvant également être prise en charge par la morphologie : {*antre pylorique*, *antre du pylore*}. Elles sont de type de celles décrites avec les règles de transformation syntaxiques implémentées par (Jacquemin, 1996). De manière générale, l'exactitude et la complétude des ressources morphologiques implémentées définissent la qualité et la couverture des relations générées.

Dans les expériences présentées, nous avons appliqué les pistes qui viennent des études linguistiques effectuées au sein de la morphologie constructionnelle. Même si les traitements automatiques conduisent aux relations en question, leur utilité dans une terminologie ne va pas de soi et doit être décidée suite à une expertise humaine et les besoins de l'application. De plus, les pistes proposées par la morphologie doivent être examinées et exploitées à côté d'autres approches. Car les apports de chacune d'elles s'avèrent souvent complémentaires.

Il est clair que l'exploitation d'indices morphologiques demande de disposer de ressources morphologiques conséquentes : un ensemble de mots reliés par tel ou tel préfixe ou suffixe, un ensemble de suffixes en cas de formation des noms d'agent et d'action. La constitution de telles ressources morphologiques constitue une tâche en soi.

Notons également, que les suffixes du paradigme relationnel peuvent acquérir un sens spécifique, par exemple lorsqu'ils apparaissent dans les noms des produits chimiques (*acide ferrique*, *acide ferreux*) : le suffixe *-ique* signifie que l'acide en question comporte un nombre d'atomes d'oxygène supérieur à celui de l'acide en *-eux* (Corbin & Paul, 1999).

7 Conclusion et perspectives

Les expériences que nous avons présentées dans cet article sont consacrées à l'étude de l'apport de la morphologie dérivationnelle pour la structuration des termes. Il apparaît que la morphologie permet de couvrir un assez large spectre de relations sémantiques. Elle contribue ainsi à la structuration de terminologies au même titre que d'autres approches existantes. Comme avec ces autres approches, l'utilisation de la morphologie demande d'une part la préparation de ressources nécessaires et d'autre part la validation des résultats. Face à la difficulté de

repérage de relations transversales entre les termes, les expériences proposées tentent d'apporter une aide et de faciliter ainsi le repérage de ces relations.

La morphologie offre deux autres modes de construction des mots: la composition et la supplétion. Toutes les deux sont potentiellement riches en relations sémantiques proposées. Leur application demande des ressources morphologiques conséquentes et des modules d'analyse et de segmentation des mots, comme celui implémenté dans (Namer & Zweigenbaum, 2004).

Références

- ALIQUOT-SUENGAS S. (1996). *Référence collective/Sens collectif. La notion du collectif à travers les noms suffixés du lexique français*. Thèse de doctorat en linguistique, Université de Lille III, Lille, France.
- AMIOT D. (1998). Préfixes ou prépositions? les cas de *sur-*, *sans-*, *contre-* et les autres. *Lexique*, **16**.
- AMSILI P. (2003). L'antonymie en terminologie: quelques remarques. In *Terminologie et Intelligence Artificielle (TIA)*, p. 31–40, Strasbourg.
- ASSADI H. (1998). *Construction d'ontologies à partir de textes techniques – Application aux systèmes documentaires*. Thèse de doctorat en informatique, Université de Paris 6, Paris, France.
- BODENREIDER O., BURGUN A. & RINDFLESCH T. C. (2001). Lexically-suggested hyponymic relations among medical terms and their representation in the UMLS. In URI INIST CNRS, Ed., *Terminologie et Intelligence artificielle (TIA)*, p. 11–21, Nancy.
- BOURIGAUT D. (1993). Analyse syntaxique locale pour le repérage de termes complexes dans un texte. *Traitement Automatique des Langues (TAL)*, p. 105–117.
- CORBIN D. (1987). *Morphologie dérivationnelle et structuration du lexique*, volume 1. Lille: Presse universitaire de Lille.
- CORBIN D. & PAUL J. (1999). Aperçu sur la créativité morphologique dans la terminologie de la chimie. *La banque des mots*, (60), 51–68.
- CÔTÉ R. A. (1996). *Répertoire d'anatomopathologie de la SNOMED internationale*, v3.4. Université de Sherbrooke, Sherbrooke, Québec.
- DAILLE B. (2003). Conceptual structuring through term variations. In *Proceedings of the ACL Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, p. 9–16.
- GARCIA D. (1997). Structuration du lexique de la causalité et réalisation d'un système d'aide au repérage de l'action dans les textes. In *Terminologie et Intelligence Artificielle (TIA)*, p. 7–26, Toulouse.
- GRABAR N. & HAMON T. (2004). Les relations dans les terminologies structurées: de la théorie à la pratique. *Revue d'Intelligence Artificielle (RIA)*, (18).
- GRABAR N. & JEANNIN B. (2002). Contribution de différents outils à la construction d'une terminologie pour la recherche d'information. In C. GRÉBOVAL, Ed., *Ingénierie des connaissances (IC)*, Rouen. Poster.
- GRABAR N. & ZWEIGENBAUM P. (2000). A general method for sifting linguistic knowledge from structured terminologies. *JAMIASUP*, p. 310–314.

- GRABAR N. & ZWEIGENBAUM P. (2003). Lexically-based terminology structuring. In *Terminology*. Sous presse.
- GREFENSTETTE G. (1994). *Explorations in automatic thesaurus discovery*. Kluwer Academic Publishers.
- HAMON T., NAZARENKO A. & GROS C. (1998). A step towards the detection of semantic variants of terms in technical documents. In *Proceedings of 17th International Conference on Computational Linguistics (COLING-ACL'98)*, p. 498–504, Université de Montréal, Montréal, Quebec, Canada.
- HEARST M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, Nantes, France. Disponible à <http://sern.ucalgary.ca/KSI/KAW/KAW99/papers/Gery1/index.html>. Visité le 26/08/99.
- JACQUEMIN C. (1996). A symbolic and surgical acquisition of terms through variation. In S. WERMTER, E. RILOFF & G. SCHELER, Eds., *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, p. 425–438, Springer.
- MANUILA L., MANUILA A., LEWALLE P. & NICOULIN M. (2001). *Dictionnaire médical*. Paris: Masson. 9^e édition.
- MÉLIS-PUCHULU A. (1991). Les adjectifs dénominaux: les adjectifs de "relation". *Lexique*, **10**, 33–60.
- MORIN E. (1999). Acquisition de patrons lexico-syntaxiques caractéristiques d'une relation sémantique. *Traitement Automatique des Langues (TAL)*, **40**(1), 143–166.
- NAMER F. & ZWEIGENBAUM P. (2004). Acquiring meaning for French medical terminology: contribution of morphosemantics. In *Annual Symposium of the American Medical Informatics Association (AMIA)*.
- NAZARENKO A., ZWEIGENBAUM P., HABERT B. & BOUAUD J. (2001). Corpus-based extension of a terminological semantic lexicon. In *Recent Advances in Computational Terminology*, p. 327–351. Benjamins.
- PACAK M. G., NORTON L. M. & DUNHAM G. S. (1980). Morphosemantic analysis of -itis forms in medical language. In *Methods in Medical Informatics (MIM)*, volume 19, p. 99–105.
- PEARSON J. (1998). *Terms in Context*, volume 1 of *Studies in Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins.
- SÉGUÉLA P. & AUSSÉNAC-GILLES N. (1999). Extraction de relations sémantiques entre termes et enrichissement de modèles du domaine. In *Actes de Ingénierie des Connaissances (IC)*, p. 79–88, Palaiseau, France.
- SPACKMAN K. & CAMPBELL K. (1998). Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies. In *Journal of American Medical Informatics Association (JAMIA)*, p. 740–744.
- ZWEIGENBAUM P. (1999). Encoder l'information médicale: des terminologies aux systèmes de représentation des connaissances. *Information de Santé Innovation Stratégie (ISIS)*, **2**(3), 27–47.